Stat 564
10-31-17

(1)

Recall <u>Generalized Least Squares</u>,

where $\mathrm{Cov}(\vec{\varepsilon}) = \sigma^2 V$

$\uparrow$ known, symmetric, positive definite

we had the solution

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

$$E[\hat{\beta}] = (X'V^{-1}X)^{-1}X'V^{-1}E[\overset{X\beta+\varepsilon}{Y}]$$

$$= (X'V^{-1}X)^{-1}X'V^{-1}X\beta = \beta$$

(2)

$$\mathrm{Cov}(\hat{\beta}) = (X'V^{-1}X)^{-1}X'V^{-1}\underbrace{\mathrm{Cov}(Y)}_{\sigma^2 V}V^{-1}X(X'V^{-1}X)^{-1}$$

$$= \sigma^2(X'V^{-1}X)^{-1}$$

Recall that $V = I \Rightarrow$ GLS is OLS

Special cases of GLS

<u>Case 1:</u> Suppose $V(\varepsilon_i) = \sigma^2 x_i$, and $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$

That is, $\mathrm{Cov}(\vec{\varepsilon}) = \sigma^2\begin{bmatrix} x_1 & & & 0 \\ & x_2 & & \\ & & \ddots & \\ 0 & & & x_n \end{bmatrix} = \sigma^2 V$

③

When V is diagonal, we refer to the solution as WLS = "weighted least squares"

$$V^{-1} = \begin{bmatrix} 1/x_1 & & 0 \\ & 1/x_2 & \\ 0 & & \ddots & \frac{1}{x_n} \end{bmatrix}$$

<u>Case II:</u> $V(\varepsilon_i) = \sigma^2 x_i^2$

$$V = \begin{bmatrix} x_1^2 & & 0 \\ & x_2^2 & \\ 0 & & \ddots & x_n^2 \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} 1/x_1^2 & & 0 \\ & 1/x_2^2 & \\ 0 & & \ddots & \frac{1}{x_n^2} \end{bmatrix}$$

④

<u>Leverage & Influence</u>

Recall $\hat{\beta} = (X'X)^{-1}X'Y$

$$\hat{Y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{H}Y$$

So $\hat{y}_i = (i^{\text{th}} \text{ row of } H) \cdot Y$

$$= \sum_{j=1}^{n} h_{ij} y_j = h_{ii} y_i + \sum_{j \neq i} h_{ij} y_j$$

That is, h_{ii} tells you how much y_i contributes to $\hat{y}_i$

Defn: h_{ii} is the leverage of the i^{th} data point

Recall $V(e_i) = \sigma^2(1-h_{ii})$

Also, $$\sum_{i=1}^{n} h_{ii} = tr \underset{n\times n}{H} = tr[\underbrace{X}_{A}\underbrace{(X'X)^{-1}X'}_{B}]$$

$$= tr[(X'X)^{-1}X'\underset{n\times p}{X}] = tr\, I_p = p$$

(6)

So the average leverage of all data points is $\frac{p}{n}$

Rule of thumb: Any data point with a leverage more than $2\cdot\frac{p}{n}$ should be examined

---

Let $\hat{\beta}$ is the least squares estimate of β

Let $\hat{\beta}_{(i)}$ be the least " " ", but with the i^{th} data point omitted

(7)

Defn: Cook's D is

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X'X (\hat{\beta}_{(i)} - \hat{\beta})}{p\, MSE}$$

D_i is the <u>influence</u> of the i^{th} data point

Rule of thumb: If Cook's $D_i > 1$, then examine that data point.

(8)

Alternate formulas for Cook's D:

$$D_i = \frac{e_i^2 h_{ii}}{(1-h_{ii})^2 p\, MSE} = \frac{r_i^2 h_{ii}}{(1-h_{ii}) p}$$

e_i := i^{th} observed residual
r_i := i^{th} studentized residual

Delta fits = DFFITS or DEFITS

$$= DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{h_{ii}}\sqrt{MSE_{(i)}}}$$

9

Alternate formula:

$$\text{DFFITS}_i = \frac{e_i\sqrt{h_{ii}}}{(1-h_{ii})\sqrt{MSE_{(i)}}}$$

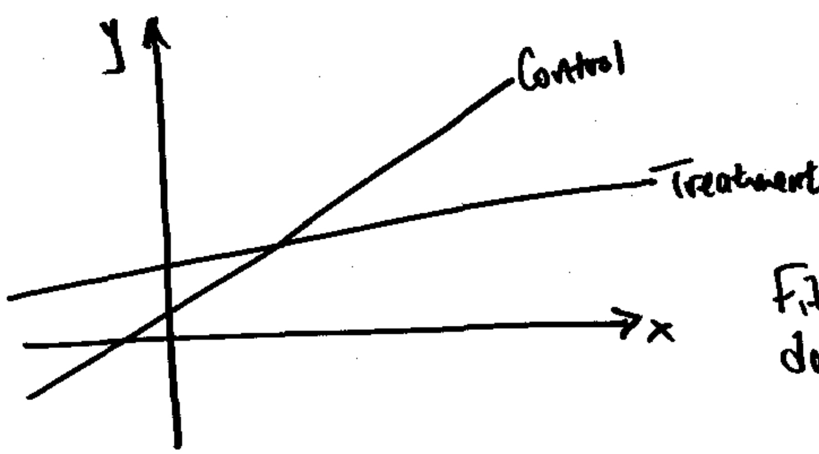
Rule of thumb: If $\text{DFFITS}_i > 2\sqrt{p/n}$, examine that data value

---

Polynomial models $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$

Use of indicators 10



Fit the entire data set in 1 model

Control: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, \ldots, m$

Treatment: $y_i = \gamma_0 + \gamma_1 x_i + \varepsilon_i \quad i = m+1, \ldots, n$

(11)

Let $W_{1i} = \begin{cases} 1 & \text{control} \\ 0 & \text{treatment} \end{cases}$

and $W_{2i} = \begin{cases} 0 & \text{control} \\ 1 & \text{treatment} \end{cases}$

$$y_i = (\beta_0 + \beta_1 x_i) W_{1i} + (\gamma_0 + \gamma_1 x_i) W_{2i} + \varepsilon_i$$

$$= \beta_0 W_{1i} + \beta_1 x_i W_{1i} + \gamma_0 W_{2i} + \gamma_1 x_i W_{2i} + \varepsilon_i$$

(12)

$$Y = X\beta + \varepsilon$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \\ \hline y_{m+1} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & 0 & 0 \\ \hline 0 & 0 & 1 & x_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{bmatrix} + \vec{\varepsilon}$$

Another solution

$$w_i = \begin{cases} 0 & \text{control} \\ 1 & \text{treatment} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_i + w_i(\delta_0 + \delta_1 x_i) + \varepsilon_i$$

$$\left[\begin{array}{c} y_1 \\ \vdots \\ y_m \\ \hline y_{m+1} \\ \vdots \\ y_n \end{array}\right] = \left[\begin{array}{cccc} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & 0 \\ \hline \vdots & \vdots & 1 & x_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & 1 & x_n \end{array}\right] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \delta_0 \\ \delta_1 \end{bmatrix} + \vec{\varepsilon}$$